

**Road Traffic Speed Prediction: A Probabilistic Model Fusing Multi-Source Data**

**Abstract:**

Road traffic speed prediction is a challenging problem in intelligent transportation system (ITS) and has gained increasing attentions. Existing works are mainly based on raw speed sensing data obtained from infrastructure sensors or probe vehicles, which, however, are limited by expensive cost of sensor deployment and maintenance. With sparse speed observations, traditional methods based only on speed sensing data are insufficient, especially when emergencies like traffic accidents occur. To address the issue, this paper aims to improve the road traffic speed prediction by fusing traditional speed sensing data with new-type “sensing” data from cross domain sources, such as tweet sensors from social media and trajectory sensors from map and traffic service platforms. Jointly modeling information from different datasets brings many challenges, including location uncertainty of low-resolution data, language ambiguity of traffic description in texts and heterogeneity of cross-domain data. In response to these challenges, we present a unified probabilistic framework, called Topic-Enhanced Gaussian Process Aggregation Model (TEGPAM), consisting of three components, i.e. location disaggregation model, traffic topic model and traffic speed Gaussian Process model, which integrate new-type data with traditional data. Experiments on real world data from two large cities in America validate the effectiveness and efficiency of our model.

**Existing System:**

Road traffic monitoring is of great importance for urban transportation system. Traffic control agencies and drivers could benefit from timely and accurate road traffic prediction and make prompt, or even advance decisions possible for detecting and avoiding road congestions. Existing methods mainly focus on raw speed sensing data collected from cameras or road sensors, and suffer severe data sparsity issue because the installation and maintenance of sensors are very expensive. At the same time, most existing techniques based only on past and

current traffic conditions do not fit well when realworld factors such as traffic accidents play a part.

**Proposed System:**

Integration of data from multiple cross-domain sources. We implement the idea of improving traffic speed prediction by integrating speed sensing data with new-type traffic-related data, such as tweet and trajectory.

Formulation of the unified TEGPAM framework. We propose a unified probabilistic framework TEGPAM that combines the disaggregation model, topic model with Gaussian Process model and is learned by variational methods and a stochastic EM algorithm.

Extensive experiments to validate the performance of the proposed method. We validate our approach using real-world data collected from two large American cities. The extensive experiments show the effectiveness of TEGPAM, as well as the model efficiency and reliability.

Elaborate analyses of introduced traffic-related data. We explore the impacts of different data sources, by decomposing TEGPAM into sub models and changing the combination ratio of datasets. Comparative experiments demonstrate the effectiveness of each data source.